PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07K 13/00, A61K 37/24, 37/26
C12P 21/06

A1

(11) International Publication Number: WO 94/07916

(43) International Publication Date: 14 April 1994 (14.04.94)

(21) International Application Number: PCT/US93/09165

(22) International Filing Date: 27 September 1993 (27.09.93)

(30) Priority data:
958,137 7 October 1992 (07.10.92) US

(60) Parent Application or Grant
(63) Related by Continuation
US 958,137 (CIP)
Filed on 7 October 1992 (07.10.92)

(71) Applicant *(for all designated States except US)*: MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : SCHMIDT, Azriel [US/US]; 54 Haymarket Lane, Bryn Mawr, PA 19010 (US). RODAN, Gideon, A. [US/US]; 827 Deerfield Lane, Bryn Mawr, PA 19010 (US). RUTLEDGE, Su, Jane [US/US]; 1886 Gerryville Pike, East Greenville, PA 18041 (US). VOGEL, Robert, L. [US/US]; 69 Holly Hill Road, Richboro, PA 18954 (US).

(74) Agent: WALLEN, John, W., III; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).

(81) Designated States: AU, BB, BG, BR, BY, CA, CZ, FI, HU, JP, KR, KZ, LK, LV, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SK, UA, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: HUMAN STEROID HORMONE RECEPTOR NERI

(57) Abstract

A novel member of the steroid hormone receptor superfamily (hereinafter identified as NERI) is disclosed which has been prepared by cDNA cloning from a human osteosarcoma SAOS-2/B10 cell library. Also disclosed is the complete sequence of human NERI complementary (DNA (Seq. ID No. 1); expression systems, including a COS stable expression system; the expressed protein (Seq. ID No. 2) and an assay using the COS expression system. NERI can be used in an assay to identify and evaluate chemical entities that bind to this receptor.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**MR**	Mauritania
AU	Australia	**GA**	Gabon	**MW**	Malawi
BB	Barbados	**GB**	United Kingdom	**NE**	Niger
BE	Belgium	**GN**	Guinea	**NL**	Netherlands
BF	Burkina Faso	**GR**	Greece	**NO**	Norway
BG	Bulgaria	**HU**	Hungary	**NZ**	New Zealand
BJ	Benin	**IE**	Ireland	**PL**	Poland
BR	Brazil	**IT**	Italy	**PT**	Portugal
BY	Belarus	**JP**	Japan	**RO**	Romania
CA	Canada	**KP**	Democratic People's Republic of Korea	**RU**	Russian Federation
CF	Central African Republic			**SD**	Sudan
CG	Congo	**KR**	Republic of Korea	**SE**	Sweden
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovak Republic
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LU**	Luxembourg	**TD**	Chad
CS	Czechoslovakia	**LV**	Latvia	**TG**	Togo
CZ	Czech Republic	**MC**	Monaco	**UA**	Ukraine
DE	Germany	**MG**	Madagascar	**US**	United States of America
DK	Denmark	**ML**	Mali	**UZ**	Uzbekistan
ES	Spain	**MN**	Mongolia	**VN**	Viet Nam
FI	Finland				

TITLE OF THE INVENTION
HUMAN STEROID HORMONE RECEPTOR NERI

SUMMARY OF THE INVENTION

The present invention relates generally to ligand-responsive regulatory proteins and genes encoding them. In particular, a novel recombinant human steroid hormone receptor (hereinafter identified as NERI) is disclosed which has been prepared by polymerase chain reaction techniques. Also disclosed is the complete sequence of human NERI complementary DNA; expression systems, including a COS stable expression system; and an assay using the COS expression system. In addition, the invention relates to a method for identifying functional ligands of the NERI receptor.

BACKGROUND OF THE INVENTION

Retinoids, steroid and thyroid hormones and possibly other molecules produce their biological effects by binding to proteins of the steroid receptor superfamily. These receptors interact with specific DNA sequences and modulate gene expression (for reviews see JM Berg, Cell 57:1065-1068 (1989); RM Evans, Science 240:899-895 (1988); M Beato, Cell 56:335-344 (1989)). Sequence analysis and functional studies of these receptors revealed two important regions which exhibit a high degree of amino acid residue conservation. The highest level of similarity among the receptors is found in a region which contains nine cystein residues that bind zinc atoms to form two "zinc fingers," which interact with the cognate steroid response elements of DNA (J Miller, et al., EMBO J 4:1609-1614 (1985); RM Evans, Cell 52:1-3 (1988)). The second region, which is less conserved, is the ligand binding domain, responsible for the interaction with the hormone (J Carlstedt-Duke, et al., Proc Natl Acad Sci USA 79:4260-4264 (1982). J. Carlstedt-Duke, et al., Proc Natl Acad Sci USA 84:4437-4440 (1987)). Recent studies have attributed additional functions to other domains of these receptors, such as protein-protein interaction that participates in transcriptional regulation (R Scule, et al., Cell

62:1217-1226 (1990); HF Yang, Cell 62:1205-1215 (1990); JM Holloway et al., Proc Natl Acad Sci USA 87:8160-8164 (1990)). The amino acid conservation in the DNA binding domain has led to the identification of new members of the steroid receptor superfamily. For example, hER1 and hER2 have been cloned by low stringency hybridization of cDNA libraries with a DNA probe coding for the DNA binding domain of the estrogen receptor (V Giguere, et al., Nature 331:91-94 (1988)). Similar approaches have led to the discovery of the retinoic acid receptors and the peroxisome proliferator activator receptor (PPAR)(I Issemann, et al., Nature 347:645-650 (1990); DJ Mangelsdorf, et al., Nature 345:224-229 (1990)). Recently, three novel members of the Xenopus nuclear hormone receptor superfamily have been disclosed (C Dreyer, Cell 68:879-887 (1992)). In addition, U.S. Patent No. 4,981,784 to Evans, et al. discloses the identification of a retinoic acid receptor and the use of chimeric constructs to produce hybrid receptors for the identification of novel ligands. The above references, however, neither disclose nor suggest the instant invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the cDNA sequence of the human NER-1 receptor (Seq. ID No. 1) and the associated expression protein (Seq. ID No. 2). The circled P indicates the amino acid proline and the boxed area represents the binding area of the protein to other DNA.

Fig. 2 illustrates in A, the three probes used for NER-1, ES 11 (Seq. ID No. 3), ES 12 (Seq. ID No. 4), and antisense ES 15 (Seq. ID No. 5). B illustrates the binding region between the ES 12 probe and the NER-1 cDNA.

DETAILED DESCRIPTION OF THE INVENTION

One embodiment of the invention concerns human steroid hormone receptor NERI, said receptor being free of other human receptor proteins.

In one class this embodiment concerns human steroid hormone receptor NERI, said receptor being free of other human proteins.

Within this class, this embodiment concerns human steroid hormone receptor NERI from human cells such as osteosarcoma, said receptor being free of other human proteins.

In a second class, this embodiment concerns a protein comprising the following 461 amino acid sequence (SEQ ID NO:2:) depicted from the amino to the carboxy terminus:

Met Ser Ser Pro Thr Thr Ser Ser Leu Asp Thr Pro Leu Pro Gly Asn
1 5 10 15

Gly Pro Pro Gln Pro Gly Ala Pro Ser Ser Ser Pro Thr Val Lys Glu
20 25 30

Glu Gly Pro Glu Pro Trp Pro Gly Gly Pro Asp Pro Asp Val Pro Gly
35 40 45

Thr Asp Glu Ala Ser Ser Ala Cys Ser Thr Asp Trp Val Ile Pro Asp
50 55 60

Pro Glu Glu Glu Pro Glu Arg Lys Arg Lys Lys Gly Pro Ala Pro Lys
65 70 75 80

Met Leu Gly His Glu Leu Cys Arg Val Cys Gly Asp Lys Ala Ser Gly
85 90 95

Phe His Tyr Asn Val Leu Ser Cys Glu Gly Cys Lys Gly Phe Phe Arg
100 105 110

Arg Ser Val Val Arg Gly Gly Ala Arg Arg Tyr Ala Cys Arg Gly Gly
115 120 125

Gly Thr Cys Gln Met Asp Ala Phe Met Arg Arg Lys Cys Gln Gln Cys
130 135 140

Arg Leu Arg Lys Cys Lys Glu Ala Gly Met Arg Glu Gln Cys Val Leu
145 150 155 160

Ser Glu Glu Gln Ile Arg Lys Lys Lys Ile Arg Lys Gln Gln Gln Gln
165 170 175

Glu Ser Gln Ser Gln Ser Gln Ser Pro Val Gly Pro Gln Gly Ser Ser
180 185 190

Ser Ser Ala Ser Gly Pro Gly Ala Ser Pro Gly Gly Ser Glu Ala Gly
195 200 205

Ser Gln Gly Ser Gly Glu Gly Glu Gly Val Gln Leu Thr Ala Ala Gln
210 215 220

Glu Leu Met Ile Gln Gln Leu Val Ala Ala Gln Leu Gln Cys Asn Lys
225 230 235 240

Arg Ser Phe Ser Asp Gln Pro Lys Val Thr Pro Trp Pro Leu Gly Ala
245 250 255

Asp Pro Gln Ser Arg Asp Ala Arg Gln Gln Arg Phe Ala His Phe Thr
260 265 270

Glu Leu Ala Ile Ile Ser Val Gln Glu Ile Val Asp Phe Ala Lys Gln
275 280 285

Val Pro Gly Phe Leu Gln Leu Gly Arg Glu Asp Gln Ile Ala Leu Leu
290 295 300

Lys Ala Ser Thr Ile Glu Ile Met Leu Leu Glu Thr Ala Arg Arg Tyr
305 310 315 320

Asn His Glu Thr Glu Cys Ile Thr Phe Leu Lys Asp Phe Thr Tyr Ser
325 330 335

Lys Asp Asp Phe His Arg Ala Gly Leu Gln Val Glu Phe Ile Asn Pro
340 345 350

Ile Phe Glu Phe Ser Arg Ala Met Arg Arg Leu Gly Leu Asp Asp Ala
355 360 365

Glu Tyr Ala Leu Leu Ile Ala Ile Asn Ile Phe Ser Ala Asp Arg Pro
370 375 380

Asn Val Gln Glu Pro Gly Arg Val Glu Ala Leu Gln Gln Pro Tyr Val
385 390 395 400

Glu Ala Leu Leu Ser Tyr Thr Arg Ile Lys Arg Pro Gln Asp Gln Leu
405 410 415

Arg Phe Pro Arg Met Leu Met Lys Leu Val Ser Leu Arg Thr Leu Ser
420 425 430

Ser Val His Ser Glu Gln Val Phe Ala Leu Arg Leu Gln Asp Lys Lys
435 440 445

Leu Pro Pro Leu Leu Ser Glu Ile Trp Asp Val His Glu.
450 455 460

or a degenerate variation thereof;

the protein being free of other human receptor proteins.

A second embodiment concerns a DNA sequence encoding human steroid hormone receptor NERI complementary DNA, said DNA, said sequence being free of other human DNA sequences.

As will be appreciated by those of skill in the art, there is a substantial amount of redundancy in the set of codons which translate specific amino acids. Accordingly, the invention also includes alternative base sequences wherein a codon (or codons) are replaced with another codon, such that the amino acid sequence translated by the DNA sequence remains unchanged. For purposes of this specification, a sequence bearing one or more such replaced codons will be defined as a degenerate variation. Also included are mutations (exchange of individual amino acids) which one of skill in the art would expect to have no effect on functionality, such as valine for leucine, arginine for lysine and asparagine for glutamine.

One class of the second embodiment of the invention concerns the following nucleotide sequence (SEQ ID NO:1:) of complementary DNA depicted from the 5' to the 3' terminus:

```
CAAGAAGTGG CGAAGTTACC TTTGAGGGTA TTTGAGTAGC GGCGGTGTGT CAGGGGCTAA  60

AGAGGAGGAC GAAGAAAAGC AGAGCAAGGG AACCCAGGGC AACAGGAGTA GTTCACTCCG 120

CGAGAGGCCG TCCACGAGAC CCCCGCGCGC AGGCATGAGC CCCGCCCCCC ACGCATGAGC 180

CCCGCCCCCC GCTGTTGCTT GGAGAGGGGC GGGACCTGGA GAGAGGCTGC TCCGTGACCC 240

CACCATGTCC TCTCCTACCA CGAGTTCCCT GGATACCCCC CTGCCTGGAA ATGGCCCCCC 300

TCAGCCTGGC GCCCCTTCTT CTTCACCCAC TGTAAAGGAG GAGGGTCCGG AGCCGTGGCC 360

CGGGGGTCCG GACCCTGATG TCCCAGGCAC TGATGAGGCC AGCTCAGCCT GCAGCACAGA 420
```

CTGGGTCATC CCAGATCCCG AAGAGGAACC AGAGCGCAAG CGAAAGAAGG GCCCAGCCCC 480

GAAGATGCTG GGCCACGAGC TTTGCCGTGT CTGTGGGGAC AAGGCCTCCG GCTTCCACTA 540

CAACGTGCTC AGCTGCGAAG GCTGCAAGGG CTTCTTCCGG CGCAGTGTGG TCCGTGGTGG 600

GGCCAGGCGC TATGCCTGCC GGGGTGGCGG AACCTGCCAG ATGGACGCTT TCATGCGGCG 660

CAAGTGCCAG CAGTGCCGGC TGCGCAAGTG CAAGGAGGCA GGGATGAGGG AGCAGTGCGT 720

CCTTTCTGAA GAACAGATCC GGAAGAAGAA GATTCGGAAA CAGCAGCAGC AGGAGTCACA 780

GTCACAGTCG CAGTCACCTG TGGGGCCGCA GGGCAGCAGC AGCTCAGCCT CTGGGCCTGG 840

GGCTTCCCCT GGTGGATCTG AGGCAGGCAG CCAGGGCTCC GGGGAAGGCG AGGGTGTCCA 900

GCTAACAGCG GCTCAAGAAC TAATGATCCA GCAGTTGGTG GCGGCCCAAC TGCAGTGCAA 960

CAAACGCTCC TTCTCCGACC AGCCCAAAGT CACGCCCTGG CCCCTGGGCG CAGACCCCCA 1020

GTCCCGAGAT GCCCGCCAGC AACGCTTTGC CCACTTCACG GAGCTGGCCA TCATCTCAGT 1080

CCAGGAGATC GTGGACTTCG CTAAGCAAGT GCCTGGTTTC CTGCAGCTGG GCCGGGAGGA 1140

CCAGATCGCC CTCCTGAAGG CATCCACTAT CGAGATCATG CTGCTAGAGA CAGCCAGGCG 1200

CTACAACCAC GAGACAGAGT GTATCACCTT CTTGAAGGAC TTCACCTACA GCAAGGACGA 1260

CTTCCACCGT GCAGGCCTGC AGGTGGAGTT CATCAACCCC ATCTTCGAGT TCTCGCGGGC 1320

CATGCGGCGG CTGGGCCTGG ACGACGCTGA GTACGCCCTG CTCATCGCCA TCAACATCTT 1380

CTCGGCCGAC CGGCCCAACG TGCAGGAGCC GGGCCGCGTG GAGGCGTTGC AGCAGCCCTA 1440

CGTGGAGGCG CTGCTGTCCT ACACGCGCAT CAAGAGGCCG CAGGACCAGC TGCGCTTCCC 1500

```
GCGCATGCTC ATGAAGCTGG TGAGCCTGCG CACGCTGAGC TCTGTGCACT CGGAGCAGGT 1560

CTTCGCCTTG CGGCTCCAGG ACAAGAAGCT GCCGCCTCTG CTGTCGGAGA TCTGGGACGT 1620

CCACGAGTGA GGGGCTGGCC ACCCAGCCCC ACAGCCTTGC CTGACCACCC TCCAGCAGAT 1680

AGACGCCGGC ACCCCTTCCT CTTCCTAGGG TGGAAGGGGC CCTGGGCGAG CCTGTAGACC 1740

TATCGGCTCT CATCCCTTGG GATAAGCCCC AGTCCAGGTC CAGGAGGCTC CCTCCCTGCC 1800

CAGCGAGTCT TCCAGAAGGG GTGAAAGGGT TGCAGGTCCC GACCACTGAC CCTTCCCGGC 1860

TGCCCTCCCT CCCCAGCTTA CACCTCAAGC CCAGCACGCA GCGTACCTTG AACAGAGGGA 1920

GGGGAGGACC CATGGCTCTC CCCCCCTAGC CCGGGAGACC AGGGGCCTTC CTCTTCCTCT 1980

GCTTTTATTT AATAAAAATA AAAACAGAAA AAAAAAAAAA AAAAAAAAAA          2030
```

A third embodiment of this invention concerns systems for expressing all or part of the human steroid hormone receptor NERI.

One class of this third embodiment of the invention comprises:

An expression construct, such as a plasmid which comprises:

a) an expression vector, such as PJ3NERI, and
b) a base sequence encoding human steroid hormone receptor NERI protein.

Within this class of the third embodiment, the steroid hormone receptor NERI comprises the nucleotide sequence (SEQ ID NO:1:) of complementary DNA as shown above.

A second class of this third embodiment of the invention concerns a system for the transient expression of human steroid

hormone receptor NERI in a suitable host cell, such as a monkey kidney cell line (COS), the system comprised of a vector which expresses human steroid hormone receptor NERI cDNA.

It is understood, and is readily apparent to those skilled in the art that a wide variety of commonly used cell lines are suitable for use in the present invention. Suitable cell lines derived from various species include, but are not limited to, cell lines of human, bovine, porcine, monkey, and rodent origin, or from yeast and bacterial strains.

A fourth embodiment of the invention concerns a method of using any of the above eukaryote or prokaryote expression systems for determining the binding affinity of a test sample for steroid hormone receptor NERI.

Following the isolation of a DNA sequence encoding human steroid hormone receptor NERI cDNA, a chimeric gene can be created by substituting the DNA-binding domain region in the DNA sequence encoding NERI cDNA with a DNA-binding domain region taken from a DNA sequence coding for another steroid hormone receptor protein, e.g., glucocorticoid (GR) receptor protein, thyroid receptor protein, mineral-ocorticoid receptor protein or retinoic acid receptor protein. Next, a suitable receptor-deficient host cell is transfected with: (1) the chimeric receptor gene, which is preferably carried on an expression plasmid, and (2) a reporter gene, such as the CAT gene or the firefly luciferase gene, which is also preferably carried on a plasmid. In any case, the reporter gene is functionally linked to an operative hormone response element (HRE) (either wild-type or engineered) wherein the hormone response element is capable of being activated by the DNA-binding domain used to make the chimeric receptor gene. (For example, if the chimeric receptor gene contains the DNA-binding domain region from glucocorticoid receptor coding DNA, then the HRE should be a wild-type, an engineered, or a synthetic GRE, i.e., one that can be activated by the operative portion of the DNA-binding region of a GR receptor protein.) Next, the transfected host cell is challenged with a test sample which contains one or more ligand(s) which can potentially bind with the ligand-binding domain region of the chimeric

protein coded for by the chimeric gene. To determine the extent that ligands can functionally complex with the chimeric receptor protein, induction of the reporter gene is monitored by monitoring changes in the protein levels of the protein coded for by the reporter gene. (For example, if luciferase is the reporter gene, the production of luciferase is indicative of receptor-regulated gene transcription.) Finally, when a ligand(s) is found that can induce transcription of the reporter gene, it is concluded that this ligand(s) can bind to the receptor protein coded for by the initial sample DNA sequence. This conclusion can be further verified by testing the binding properties of the receptor protein, coded for by the initial sample DNA sequences, vis-a-vis the ligand(s) that induce expression of the reporter gene.

The fourth embodiment further concerns a method for determining the affinity of a test sample for activation of a steroid hormone receptor NERI, the method comprising:

(a) constructing a chimeric gene by substituting portions of a DNA-binding domain region of a DNA sequence encoding human steroid hormone receptor NERI cDNA with operative portions of a DNA-binding domain region from a known ligand-responsive receptor protein;

(b) introducing into a suitable receptor-deficient host cell:

(i) the chimeric gene from step (a), and

(ii) a reporter gene functionally linked to an operative hormone response element wherein the hormone response element is capable of being activated by the DNA-binding domain region of the receptor protein encoded by the chimeric gene of step (a);

(c) challenging the transfected host cell from step (b) with the test sample to be evaluated for ligand-binding activity with the chimeric receptor protein encoded by the chimeric gene of step (a);

(d) assaying induction of the reporter gene by monitoring changes in the protein levels of the protein coded for by the reporter gene.

One class of this embodiment concerns a method of using a monkey kidney cell line (COS) as the suitable receptor-deficient host cell. In addition the COS host cell line may be transfected with a plasmid, the plasmid comprising:

(a) an expression vector, such as PJ3NERI, and

(b) the base sequence encoding human steroid hormone receptor NERI protein.

The aforementioned fourth embodiment is further useful for identifying compounds which may be peroxisome proliferators and, hence, are potentially hepatocarcinogens. This embodiment is also useful in identifying ligands for new hormone systems which regulate bodily function.

In overview, the present invention describes methods to isolate the human steroid hormone receptor NERI complementary DNA (cDNA) without prior knowledge of its protein sequence or gene sequence. Polymerase chain reaction (PCR) technique was utilized for the isolation of human steroid hormone receptor NERI cDNA.

The complete sequence of the human steroid hormone receptor NERI cDNA was determined, and its encoded protein sequence was deduced. Among other things, such sequence information is useful in the process of developing novel steroid hormone agonists and antagonists.

An expression system was used to express the cloned human steroid hormone receptor NERI cDNA. The COS (a monkey kidney cell line) expression system can be used to measure the ligand binding properties of human steroid hormone receptor NERI.

Assay protocols use the heterologously expressed human steroid hormone receptor NERI for determination of the activation of steroid hormone receptor NERI by antagonists.

The present invention generally relates to a new member of the steroid hormone receptor superfamily. The amino acid sequence deduced from the DNA sequence (Bases 245 to 1027) shows the characteristic features of both the DNA and the ligand binding domains of this family of receptors. Sequence analysis predicted a protein of

461 amino acids which includes the conserved amino acid residues characteristic of the DNA and ligand-binding domains of nuclear receptors.

This invention relates to Ner-I, a new member of the steroid receptor-like gene family which was isolated from a human bone cell cDNA library. Ner-I codes for a polypeptide of 461 amino acids which contains the conserved sequences of the DNA and ligand binding domains of typical steroid receptors. The best homology is shared with the different retinoic acid receptors: α, ß & γ; 55% at the DNA α, γ binding domain and 38-40% at the ligand binding domain. A single transcript of 2.3kb was detected in all cells and tissues tested. We tested the potential of these constructs to mediate ligand dependent transcription activation of reporter genes. To date, no specific ligand for this receptor was identified but it is reasonably believed that binding will occur with a member of a retinoic acid receptor family.

The nuclear receptor-gene family is expanding in size, as new members are constantly identified. Here we report the cloning of a new sequence from human osteosarcoma cells. This gene, named Ner-I, codes for a polypeptide of 461 amino acids and contains the conserved sequences typical of both the DNA and the ligand binding domains. The amino terminal of the predicted protein contains a high number of proline and serine residues which might introduce a highly stabilized and complexed secondary structure. A high number of proline residues was also found in other nuclear receptor and other molecules with transcriptional activity such as CTf/N1, fos, jun. p53, OCT-2 and SRF (Mitchell & Tjian, Science, 245, pp. 371-378 (1989); Mermod et al 1989).

The size of the deduced protein and the spatial distribution of the different domains resemble the arrangement found in the thyroid, vitamin D and retinoic acid-receptor subgroup (Lazar et al., Proc Natl. Acad. Sci. 86, pp. 7771-7774, 1989). The sequence homology at the predicted ligand binding domain ranges between 33-40% identity with the members of this subgroup, while homologies lower than 25% were measured when the ligand binding domain was compared to the

corresponding domain of the steroid receptor subgroup which includes the estrogen, glucocorticoid, androgen and progesterone. As mentioned above the highest homology of the ligand binding domain was the retinoic acid receptors. This homology, 40% with retinoic acid type A is much lower than homologies of 79% and over which are found between RARα, RARß, RARγ. The degree of sequence similarity however, is not always indicative of the nature of the ligand as evident from the recently discovered new form of retinoic acid receptor, RXR which shared only a 27% identity with the other retinoic acid receptors (Oro et al, Nature, 347, pp. 298-301, 1990). It is thus impossible to assign or to exclude any of the known ligand based on sequence homology considerations. The homology at the DNA binding domain is around 50% with most other nuclear receptors. The highest degrees of homology were measured with estrogen and with retinoic acid receptors, 56%, and 53-55% respectively. However, these levels were only marginally higher than the homologies with the other receptors. It is worth noting that the homology shared between the different retinoic acid receptors (types A, ß and gamma) at this domain are higher than 95%. And even the homology of RXR to the other retinoic acid receptors at this region excedes 60%.

Although cloned from an osteoblastic cell line, the mRNA for Ner-I is widely distributed in different tissues and in all the tested cell lines.

To simplify the search for the elusive ligand, we constructed a hybrid receptor gene comprising the DNA binding domain of estrogen receptor linked to the ligand binding domain of the Ner-I gene. Such strategy was proven successful in the identification of ligands for the PPAR receptor Issenmann and Green, Nature, 347, pp. 645-649 (1990). We, however, failed to identify specific induction with any of the ligand tested. This result should not come as a surprise since the number of potential candidates is very large. Molecules which are very remote in structure from classical steroids or vitamins may serve as ligand for the receptor as evident from the identification of the PPAR as a member of this family. The ligands for such receptors might also

be intracellular factors which interacts and modulates the transcriptional activity of other receptor-like molecule.

In summary, we have identified a new member of the steroid hormone receptor superfamily. The identification of these functions may provide us with an insight into a novel hormonal regulated system.

As used herein, "steroid hormone receptor superfamily" refers to the class of related receptors comprised of glucocorticoid, mineralocorticoid, progesterone, estrogen, estrogen-related, vitamin D_3, thyroid, v-erb-A, retinoic acid and E75 (Drosophilia) receptors. As used herein "steroid hormone receptor" refers to members within the steroid hormone receptor superfamily.

As used herein, "ligand" means an inducer, such as a hormone or growth substance. Inside a cell the ligand binds to a receptor protein, thereby creating a ligand-receptor complex, which in turn can bind to an appropriate hormone response element. Single ligands may have multiple receptors.

As used herein, "expression construct" refers to a plasmid or vector comprising a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. "Recombinant expression system" means a combination of an expression construct and a suitable host microorganism.

The following examples are given for the purpose of illustrating the present invention and shall not be construed as being limitations on the scope or spirit of the instant invention.

EXAMPLE 1

Primers design

Degenerate DNA primers were designed to recognize the consensus sequences of the DNA and ligand binding domains of a typical

nuclear receptor. The 5' primer ES11, (Seq. ID. No. 3) was degenerate oligomer 5' TGTGAGGGCTGCAA(G/A)G(C/G)C, based on the conserved amino acids CEGCKA(G) of the DNA binding domain. A second 5' primer, ES12, (Seq. ID. No. 4) TGTGAGGGCTGCAA(G/A)G(C/G)CTTCTTC contains six additional nucleodites at its 3'-end corresponding to two conserved phenyalanine residues following the CEGCKA(G) sequence. The antisense primer, ES15 (Seq. ID. No. 5) AA(G)A(C,T,G)CCA(C,T,G)GGIAIIIIC(T)TTT(A,G,C)GC(G)TT, was designed to complement the semiconserved aminoacid sequence FAKxxPGF of the ligand binding domain of a typical receptor. The nucleotides corresponding to the nonconserved aminoacids (XX) were substituted with inosine (I) residues.

PCR Amplification

To use the polymerase chain reaction (PCR) method, degenerate oligonucleotides were synthesized according to the amino acid sequence of two conserved segments shared by members of the nuclear receptor superfamily (RM Evans, Science 240:899-895 (1988)). The 5'end primers, ES11 and ES 12, were designed according to a segment of the DNA binding domain. The primer at the 3' end, ES 15, was prepared according to a conserved amino acid sequence in the ligand binding domain of the retinoid receptor subfamily and the vitamin D receptor. Since this conserved region contains two nonconserved amino acid residues, inosine nucleotides were used as part of this primer. Human cDNA prepared from mRNA of osteosarcoma cells SAOS-2/B10, amplified with the primers ES11 and ES15, yielded multiple DNA fragments with various sizes after the first round of amplification. A portion of the reaction was subjected to a second round of amplification using the nested primer ES12 and the same 3' end primer ES15.

A random primed cDNA library was prepared from 2 μg total RNA isolated from the osteosarcoma SAOS-2/B10 cells by the Moloney reverse transcriptase enzyme RTH according to the

manufacturer recommendations (Bethesda Research Laboratories). The cDNA reaction (25 µl) was diluted into 300 ml water and heat denatured at 95°C for 5 minutes and quickly chilled on ice. The cDNA (2.5 µl) and the first primer pair, ES11 and ES15 (0.5 µM each) were employed in the amplification reaction with the amplitaq kit and the DNA thermal cycler (Perkin-Elmer-Cetus).

Primer ES11 has the following sequence (SEQ ID NO:3:):

CGAATTCTGT GAGGGCTGGA ARGSC 25

wherein: R represents A or G; and
S represents C or G;

and Primer ES15 has the following sequence (SEQ ID NO:5:):

GGAATTCRAA NCCNGGNANN NNYTTNGCRA A 31

wherein: N (at the 11, 14 & 26 positions) represents A or C or G or T; N (at the 17, 19, 20, 21 & 22 positions) represent inosine; R represents A or G; S represents C or G; and Y represents C or T.

The following amplification cycles were conducted: denaturation at 94°C, 1.5 minutes; annealing at 65°C, 3 minutes; extension at 72°C, 5 minutes for 3 cycles; denaturing at 94°C, 1 minute; annealing at 60°C, 3 minutes; extension at 72°C, 5 minutes for 15 cycles; and denaturing at 94°C, 1 minute; annealing at 57°C, 3 minutes; extension at 72°C, 5 minutes for 20 cycles.

After completion of the first round of amplification, 5 µl of the reaction was added to an amplification reaction buffer containing a second set of primers: a partially nested oligomer ES12 and the same 3' end primer ES15 (0.5 µM each).

Primer ES12 has the following sequence (SEQ ID NO:4:):

CGAATTCTGT GAGGGCTGCA ARGSCTTCTT C 31

wherein: R represents A or G; and
S represents C or G.

The second round of amplification was performed with the same program used for the first amplification cycles. The amplification products were separated on 5% polyacrylamide gel and stained by ethidium bromide. The DNA products were isolated from the gel, phosphorylated by T4 polynucleotide kinase and cloned into PUC 18 vector by blunt end ligation. Clones were identified by digestion of plasmid DNA with PvuII enzyme. The DNA insert was analyzed by double-stranded DNA sequencing by the dideoxy termination method using sequenase enzyme kit (United States Biochemicals).

This amplification produced two major DNA fragments of 270 bp and 320, respectively.

cDNA Amplification

Single stranded randomly primed cDNA was prepared with the Mo-MLV reverse transcriptase (BRL) from RNA isolated from Saos-2/B10, a human osteosarcoma cell line (Rodan et al., Cancer Research, 47], pp. 4961-4966, 1987; Endocrinol, 122, pp. 219-227, (1989). The cDNA reaction (25 µl) was diluted into 300 µl water and heat denaturated at 95°C for 5 minutes and quickly chilled on ice. The cDNA (2.5 µl) and the first primer pair, ES11 and ES15 (0.5 µM each), were employed in the amplification reaction with the amplitaq kit and the DNA thermal cycler (Perkin Elmer, Cetus). We carried out the following amplification cycles; denaturation at 94°C, 1.5 minutes; annealing at 65°C 3 minutes; extension at 72°C, 5 minutes for 3 cycles; denaturation at 94°C, 1 minute; annealing at 60°C, 3 minutes; extension at 72°C, 5 minutes for 15 cycles; and denaturating at 94°C, 1 minute; annealing at 57°C, 3 minutes; extension at 72°C, 5 minutes for 20 cycles. After completion of the first round of amplification, 5% of the reaction were added to an amplification reaction buffer containing a second set of primers: a partially nested oligomer ES12 and the same 3'-end primer, ES15 (0.5 µM each). The second round of amplification was performed with the same program used for the first amplification

cycles. The amplified fragments were separated by electrophoresis on 5% polyacrylamide gel, cloned into plasmids and sequenced. The DNA fragments were then used to screen a lamda gt11 cDNA library of Saos-2/B10 cells. Positive clones were isolated and sequenced in a bidirectional way by the sequence kit (United States Biochemicals).

EXAMPLE 2

Cloning and Sequencing of cDNA

A human oligo-dT cDNA library was constructed RNA isolated from osteosarcoma SAOS-2/B10 cells using the Lambda Librarian cloning kit (Invitrogen Corp.). Several positive clones were identified by plaque screening with the [32P] labeled DNA probe of the cloned amplified product (NERI). The hybridization conditions were as described by A Schmidt, et al., J Biol Chem 259:7411-7415 (1984). The cDNA inserts were cloned into EcoRI site of the cloning vector PUC18. The complete DNA sequence of both strands was determined by the dideoxy sequencing method using a series of oligonucleotides synthesized as the DNA sequence data became available.

The fragments from PCR amplification were cloned into plasmids and sequenced. The amino acid residues predicted by the DNA sequences, indicated that both DNA fragments may code for genuine and novel receptors belonging to the steroid hormone superfamily. To obtain the complete cDNA clone the amplified cDNA fragment of 270 bp NERI was used for the screening of a human osteosarcoma SAOS-2/B10 cells cDNA library. All the highly positive clones were identical and matched the sequence for the amplified NERI DNA fragment.

EXAMPLE 3

Northern Blot Analysis

RNA from various tissues or the listed cell lines were prepared by using guanidine thiocyanate or by the guanidine hydrochloride method (GGA Nemeth, et al., Anal Biochem 183:301-

304 (1989); JM Chirgwin, et al., Biochemistry, 18:5294-5299 (1979)). RNA samples were analyzed by formaldehyde agarose gel electrophoresis as described by (KM Rosen, et al., Focus 12:23-24 (1990)). The RNA was transferred by blotting to N-Hybond (Amersham Corp.), and hybridized with ^{32}P-labeled cDNA of NERI as described by (A Schmidt, et al., J Biol Chem 259:7411-7415 (1984); KM Rosen, et al., Focus 12:23-24 (1990)).

Expression of NERI mRNA

Total RNA was extracted from rat or baboon tissues and processed for electrophoresis and blot hybridization with ^{32}p labeled probe of Ner-I by conventional methods as described by Fritsch et al (1989).

RESULTS

Cloning of Ner-I

The amplification of the cDNA prepared from the RNA of Saos-2/B10 osteoblastic cell line with the ES11 and ES15 primers yielded multiple fragments after 40 rounds of amplification. Five percent of the first amplification reaction were subjected to additional 30 rounds of amplication with ES12 and ES15 oligomers. Primer ES12 that replaces ES11 is six nucleodites longer and codes for two conserved phenylalanine residues at the 3'-end, thus introduces an additional level of specificity to the amplification reaction. The second amplification step resulted in the elimination of all but two DNA fragments. The two fragments; nuc-1, 320bp and Ner-I, 270bp, were subcloned and sequenced. Sequence analysis revealed that both DNA fragments resemble the typical DNA binding domain of steroid hormone receptor genes, but were not identical to any of the known sequences.

Surprisingly, none of the two fragments contained sequences of the ligand binding domain as could be predicted by the use of the ES15 primer derived from that region. It was later realized that

the 5' ES12 oligomer primed the reaction at both directions although it shared only 53% homology with that sequence.

In order to obtain full length cDNA clone for the novel putative nuclear receptor Ner-I, we screened a cDNA library from the Saos-2/B10 cells with the Ner-I amplified DNA fragment. Several clones were identified and cloned into pUC18 vectors. One of the largest clones, nuc-2-103 of 2kb was thoroughly analyzed and the nucleodite sequence and the predicted amino acid sequence were determined.

Sequencing of the Ner-I clone revealed a long open reading frame coding for a polypeptide of 461 amino acids. The deduced protein resembles in its structure a typical steroid-like receptor. At position 87-154, a putative "double zinc finger" structure which can serve as the DNA binding domain was identified. Amino acid sequences that characterize the ligand binding domain were located toward the carboxy terminus of the protein and were spaced like in the thyroid or retinoic acid receptors. Comparing the sequence of the deduced protein with other known receptor sequences revealed that the DNA binding domain shared 50-56% homology with all the steroid-like receptors. Highest scores at this domain were: 56% for the estrogen receptor, 55% for the retinoic acid gamma receptor and mineralcorticoid receptor and 54% for retinoic acid A and glucocorticoid receptors. The ligand binding domain which is less conserved showed highest homology levels of 38-40% with the 3 types of retinoic acid receptors, RARα, RARß and RARγ 38% with vitamin D receptor and 33% with thyroid hormone receptor. The homology to the ligand binding domains of estrogen, androgen, glucocorticoid and mineralocorticoid at this domain was significantly lower. The RXR retinoic acid receptor type X showed an intermediate value of 28% homology at this domain.

It is noteworthy that the amino acid terminus of Ner-I (amino acids 1-87) contains high number of 17 proline residues and 10 serines.

Expression of Ner-I

Analysis of RNA from the osteoblastic Saos-2/B10 cells with the Ner-I labeled DNA probe revealed a single transcript of approximate 2.3 kb. Similar RNA transcripts were detected in all cell lines tested. No apparent variations in size of the mRNA molecules could be observed between RNAs isolated from different species. Tissue distribution of the Ner-I gene expression was examined by Northern hybridization. Ner-I RNA transcripts were detected in all the rat tissues which were tested. Similar results were obtained with RNA isolated from tissues of adult baboons.

Studying the transcriptional activation properties of ner-I

In order to try to identify the putative ligand for Ner-I receptor, we set on examining the potential of Ner-I to induce transcription of a reporter gene which contains inducible hormone responsive elements. Several responsive elements were tested; the thyroid/retinoic acid, estrogen, vitamin D and the glucocorticoid/ progesterone elements. Transfection experiments in cv-1 and L cells, revealed no ligand-dependent induction of the CAT reporter gene. To facilitate the search for a ligand, we have constructed hybrid receptor molecules. The DNA region coding for the putative ligand binding domain of Ner-I was fused in frame to the region coding for the amino terminal and DNA binding domain of the estrogen or glucocorticoid (GR) receptors. This recombinant receptor was employed in ligand transcription experiments using the pERE-BLCAT plasmid as a reporter gene (Lukow and Schultz, Nuc. Acid Res. 15, pp. 5490-5491, (1987) or MMTV-luciferase reporter plasmid. We tested several metabolites of vitamin A, Vitamin D, Vitamin E, thyroid hormone, estrogen, progesterone and other potential ligands for their activity in this assay, but could not detect specific response with any of the ligands. However, it is reasonably believed that a steroid hormone exists for which NER-I binds to and associated with biological activity at concentrations of 1-10 micromolar.

Northern analysis with a NERI cDNA probe revealed that NERI receptor mRNA is expressed as a 2.3 Kb transcript in the human osteosarcoma SAOS-2/B10 cells.

Similar RNA transcripts were identified in other cell lines and tissues that were tested.

Screening the Saos-2/b10 cDNA library with the labeled amplified DNA fragment encoding part of the putative novel nuclear receptor NER I resulted in several positive cDNA clones. Sequence analysis of the positive clones revealed that in addition to the expected full length cDNA clone for NER I receptor we obtained two clones in which the DNA sequences differed from the expected NER I putative receptor. The sequence of one clone, named pE1001, matched the sequence of the known retinoic acid receptor type alpha (RARα), (Giguere et al., Nature 331, pp. 91-94, 1987). Sequence analysis of the second clone (pE1005), revealed the characteristics of a novel nuclear receptor published and characterized as a novel retinoic acid receptor X, (RXRα) (Mangelsdorf et al., Nature 345, pp. 224-229 1990). Thus, these results illustrate that the cDNA for NER I receptor can be utilized as an assay tool to identify known and novel members of the class of steroid hormone nuclear receptors.

While the foregoing specifications teaches the principles of the present invention, with examples provided for the purpose of illustration, it will be understood that the practice of the invention encompasses all of the casual variations, adaptations, modifications, deletions, or additions of procedures and protocols described herein, as come within the scope of the following claims and it equivalents.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Schmidt, Azriel
Rutledge, Su Jane
Shinar, Doron
Rodan, Gideon

(ii) TITLE OF INVENTION: Human Steroid Hormone Receptor NER1

(iii) NUMBER OF SEQUENCES: 2

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: John W. Wallen III
(B) STREET: 126 East Lincoln Avenue
(C) CITY: Rahway
(D) STATE: New Jersey
(E) COUNTRY: US
(F) ZIP: 07065-0907

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: Macintosh IIci
(C) OPERATING SYSTEM: System 7.0.1
(D) SOFTWARE: Microsoft Word 5.0a

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Wallen, John W. III
(B) REGISTRATION NUMBER: 35,403
(C) REFERENCE/DOCKET NUMBER: 18686

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (908) 594-3905
(B) TELEFAX: (908) 574-4720

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1979 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CAAGAAGTGG CGAAGTTACC TTTGAGGGTA TTTGAGTAGC GGCGGTGTGT CAGGGGCTAA 60

AGAGGAGGAC GAAGAAAAGC AGAGCAAGGG AACCCAGGGC AACAGGAGTA GTTCACTCCG 120

CGAGAGGCCG TCCACGAGAC CCCCGCGCGC AGGCATGAGC CCCGCCCCCC ACGCATGAGC 180

CCCGCCCCCC GCTGTTGCTT GGAGAGgGGC GGGACCTGGA GAGAGGCTGC TCCGTGACCC 240

CACCATGTCC TCTCCTACCA CGAGTTCCCT GGATACCCCC CTGCCTGGAA ATGGCCCCCC 300

TCAGCCTGGC GCCCCTTCTT CTTCACCCAC TGTAAAGGAG GAGGGTCCGG AGCCGTGGCC 360

CGGGGGTCCG GACCCTGATG TCCCAGGCAC TGATGAGGCC AGCTCAGCCT GCAGCACAGA 420

CTGGGTCATC CCAGATCCCG AAGAGGAACC AGAGCGCAAG CGAAAGAAGG GCCCAGCCCC 480

GAAGATGCTG GGCCACGAGC TTTGCCGTGT CTGTGGGGAC AAGGCCTCCG GCTTCCACTA 540

CAACGTGCTC AGCTGCGAAG GCTGCAAGGG CTTCTTCCGG CGCAGTGTGG TCCGTGGTGG 600

GGCCAGGCGC TATGCCTGCC GGGGTGGCGG AACCTGCCAG ATGGACGCTT TCATGCGGCG 660

CAAGTGCCAG CAGTGCCGGC TGCGCAAGTG CAAGGAGGCA GGGATGAGGG AGCAGTGCGT 720

CCTTTCTGAA GAACAGATCC GGAAGAAGAA GATTCGGAAA CAGCAGCAGC AGGAGTCACA 780

GTCACAGTCG CAGTCACCTG TGGGGCCGCA GGGCAGCAGC AGCTCAGCCT CTGGGCCTGG 840

GGCTTCCCCT GGTGGATCTG AGGCAGGCAG CCAGGGCTCC GGGGAAGGCG AGGGTGTCCA 900

GCTAACAGCG GCTCAAGAAC TAATGATCCA GCAGTTGGTG GCGGCCCAAC TGCAGTGCAA 960

CAAACGCTCC TTCTCCGACC AGCCCAAAGT CACGCCCTGG CCCCTGGGCG CAGACCCCCA 1020

GTCCCGAGAT GCCCGCCAGC AACGCTTTGC CCACTTCACG GAGCTGGCCA TCATCTCAGT 1080

CCAGGAGATC GTGGACTTCG CTAAGCAAGT GCCTGGTTTC CTGCAGCTGG GCCGGGAGGA 1140

CCAGATCGCC CTCCTGAAGG CATCCACTAT CGAGATCATG CTGCTAGAGA CAGCCAGGCG 1200

CTACAACCAC GAGACAGAGT GTATCACCTT CTTGAAGGAC TTCACCTACA GCAAGGACGA 1260

CTTCCACCGT GCAGGCCTGC AGGTGGAGTT CATCAACCCC ATCTTCGAGT TCTCGCGGGC 1320

CATGCGGCGG CTGGGCCTGG ACGACGCTGA GTACGCCCTG CTCATCGCCA TCAACATCTT 1380

CTCGGCCGAC CGGCCCAACG TGCAGGAGCC GGGCCGCGTG GAGGCGTTGC AGCAGCCCTA 1440

CGTGGAGGCG CTGCTGTCCT ACACGCGCAT CAAGAGGCCG CAGGACCAGC TGCGCTTCCC 1500

GCGCATGCTC ATGAAGCTGG TGAGCCTGCG CACGCTGAGC TCTGTGCACT CGGAGCAGGT 1560

CTTCGCCTTG CGGCTCCAGG ACAAGAAGCT GCCGCCTCTG CTGTCGGAGA TCTGGGACGT 1620

CCACGAGTGA GGGGCTGGCC ACCCAGCCCC ACAGCCTTGC CTGACCACCC TCCAGCAGAT 1680

AGACGCCGGC ACCCCTTCCT CTTCCTAGGG TGGAAGGGGC CCTGGGCGAG CCTGTAGACC 1740

TATCGGCTCT CATCCCTTGG GATAAGCCCC AGTCCAGGTC CAGGAGGCTC CCTCCCTGCC 1800

CAGCGAGTCT TCCAGAAGGG GTGAAAGGGT TGCAGGTCCC GACCACTGAC CCTTCCCGGC 1860

TGCCCTCCCT CCCCAGCTTA CACCTCAAGC CCAGCACGCA GCGTACCTTG AACAGAGGGA 1920

GGGGAGGACC CATGGCTCTC CCCCCCTAGC CCGGGAGACC AGGGGCCTTC CTCTTCCTCT 1980

GCTTTTATTT AATAAAAATA AAAACAGAAA AAAAAAAAAA AAAAAAAAAA 2030

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 454 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ser Ser Pro Thr Thr Ser Ser Leu Asp Thr Pro Leu Pro Gly Asn
1               5                  10                  15

Gly Pro Pro Gln Pro Gly Ala Pro Ser Ser Ser Pro Thr Val Lys Glu
            20                  25                  30

Glu Gly Pro Glu Pro Trp Pro Gly Gly Pro Asp Pro Asp Val Pro Gly
        35                  40                  45

Thr Asp Glu Ala Ser Ser Ala Cys Ser Thr Asp Trp Val Ile Pro Asp
    50                  55                  60

Pro Glu Glu Glu Pro Glu Arg Lys Arg Lys Lys Gly Pro Ala Pro Lys
65                  70                  75                  80

Met Leu Gly His Glu Leu Cys Arg Val Cys Gly Asp Lys Ala Ser Gly
                85                  90                  95
```

```
Phe His Tyr Asn Val Leu Ser Cys Glu Gly Cys Lys Gly Phe Phe Arg
            100                 105                 110

Arg Ser Val Val Arg Gly Gly Ala Arg Arg Tyr Ala Cys Arg Gly Gly
        115                 120                 125

Gly Thr Cys Gln Met Asp Ala Phe Met Arg Arg Lys Cys Gln Gln Cys
    130                 135                 140

Arg Leu Arg Lys Cys Lys Glu Ala Gly Met Arg Glu Gln Cys Val Leu
145                 150                 155                 160

Ser Glu Glu Gln Ile Arg Lys Lys Lys Ile Arg Lys Gln Gln Gln Gln
                165                 170                 175

Glu Ser Gln Ser Gln Ser Gln Ser Pro Val Gly Pro Gln Gly Ser Ser
            180                 185                 190

Ser Ser Ala Ser Gly Pro Gly Ala Ser Pro Gly Gly Ser Glu Ala Gly
        195                 200                 205

Ser Gln Gly Ser Gly Glu Gly Glu Gly Val Gln Leu Thr Ala Ala Gln
    210                 215                 220

Glu Leu Met Ile Gln Gln Leu Val Ala Ala Gln Leu Gln Cys Asn Lys
225                 230                 235                 240

Arg Ser Phe Ser Asp Gln Pro Lys Val Thr Pro Trp Pro Leu Gly Ala
                245                 250                 255

Asp Pro Gln Ser Arg Asp Ala Arg Gln Gln Arg Phe Ala His Phe Thr
            260                 265                 270

Glu Leu Ala Ile Ile Ser Val Gln Glu Ile Val Asp Phe Ala Lys Gln
        275                 280                 285

Val Pro Gly Phe Leu Gln Leu Gly Arg Glu Asp Gln Ile Ala Leu Leu
    290                 295                 300

Lys Ala Ser Thr Ile Glu Ile Met Leu Leu Glu Thr Ala Arg Arg Tyr
305                 310                 315                 320

Asn His Glu Thr Glu Cys Ile Thr Phe Leu Lys Asp Phe Thr Tyr Ser
                325                 330                 335

Lys Asp Asp Phe His Arg Ala Gly Leu Gln Val Glu Phe Ile Asn Pro
            340                 345                 350

Ile Phe Glu Phe Ser Arg Ala Met Arg Arg Leu Gly Leu Asp Asp Ala
        355                 360                 365

Glu Tyr Ala Leu Leu Ile Ala Ile Asn Ile Phe Ser Ala Asp Arg Pro
    370                 375                 380

Asn Val Gln Glu Pro Gly Arg Val Glu Ala Leu Gln Gln Pro Tyr Val
385                 390                 395                 400
```

Glu Ala Leu Leu Ser Tyr Thr Arg Ile Lys Arg Pro Gln Asp Gln Leu
405 410 415

Arg Phe Pro Arg Met Leu Met Lys Leu Val Ser Leu Arg Thr Leu Ser
420 425 430

Ser Val His Ser Glu Gln Val Phe Ala Leu Arg Leu Gln Asp Lys Lys
435 440 445

Leu Pro Pro Leu Leu Ser Glu Ile Trp Asp Val His Glu
450 455 460

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CGAATTCTGT GAGGGCTGGA ARGSC 25

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGAATTCTGT GAGGGCTGCA ARGSCTTCTTC 31

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGAATTCRAA NCCNGGNANN NNYTTNGCRA A
31

WHAT IS CLAIMED IS:

1. A human steroid receptor NERI, the receptor being substantially free of other human receptor proteins.

2. The human steroid receptor NERI of Claim 1, the receptor being free of other human proteins.

3. The human steroid receptor NERI of Claim 2, the receptor being a recombinantly produced receptor from human cells.

4. A protein corresponding to the amino acid sequence of human steroid receptor NERI, the protein comprising 461 amino acids.

5. The protein of Claim 4 comprising the amino acid sequence (SEQ ID NO:2:) which is:

6. A DNA sequence encoding human steroid receptor NERI, the sequence being free of other human DNA sequences.

7. The DNA sequence of Claim 6 comprising the sequence (SEQ ID NO:1:) which is:

```
CAAGAAGTGG CGAAGTTACC TTTGAGGGTA TTTGAGTAGC GGCGGTGTGT CAGGGGCTAA 60

AGAGGAGGAC GAAGAAAAGC AGAGCAAGGG AACCCAGGGC AACAGGAGTA GTTCACTCCG 120

CGAGAGGCCG TCCACGAGAC CCCCGCGCGC AGGCATGAGC CCCGCCCCCC ACGCATGAGC 180

CCCGCCCCCC GCTGTTGCTT GGAGAGgGGC GGGACCTGGA GAGAGGCTGC TCCGTGACCC 240

CACCATGTCC TCTCCTACCA CGAGTTCCCT GGATACCCCC CTGCCTGGAA ATGGCCCCCC 300
```

TCAGCCTGGC GCCCCTTCTT CTTCACCCAC TGTAAAGGAG GAGGGTCCGG AGCCGTGGCC 360

CGGGGGTCCG GACCCTGATG TCCCAGGCAC TGATGAGGCC AGCTCAGCCT GCAGCACAGA 420

CTGGGTCATC CCAGATCCCG AAGAGGAACC AGAGCGCAAG CGAAAGAAGG GCCCAGCCCC 480

GAAGATGCTG GGCCACGAGC TTTGCCGTGT CTGTGGGGAC AAGGCCTCCG GCTTCCACTA 540

CAACGTGCTC AGCTGCGAAG GCTGCAAGGG CTTCTTCCGG CGCAGTGTGG TCCGTGGTGG 600

GGCCAGGCGC TATGCCTGCC GGGGTGGCGG AACCTGCCAG ATGGACGCTT TCATGCGGCG 660

CAAGTGCCAG CAGTGCCGGC TGCGCAAGTG CAAGGAGGCA GGGATGAGGG AGCAGTGCGT 720

CCTTTCTGAA GAACAGATCC GGAAGAAGAA GATTCGGAAA CAGCAGCAGC AGGAGTCACA 780

GTCACAGTCG CAGTCACCTG TGGGGCCGCA GGGCAGCAGC AGCTCAGCCT CTGGGCCTGG 840

GGCTTCCCCT GGTGGATCTG AGGCAGGCAG CCAGGGCTCC GGGGAAGGCG AGGGTGTCCA 900

GCTAACAGCG GCTCAAGAAC TAATGATCCA GCAGTTGGTG GCGGCCCAAC TGCAGTGCAA 960

CAAACGCTCC TTCTCCGACC AGCCCAAAGT CACGCCCTGG CCCCTGGGCG CAGACCCCCA 1020

GTCCCGAGAT GCCCGCCAGC AACGCTTTGC CCACTTCACG GAGCTGGCCA TCATCTCAGT 1080

CCAGGAGATC GTGGACTTCG CTAAGCAAGT GCCTGGTTTC CTGCAGCTGG GCCGGGAGGA 1140

CCAGATCGCC CTCCTGAAGG CATCCACTAT CGAGATCATG CTGCTAGAGA CAGCCAGGCG 1200

CTACAACCAC GAGACAGAGT GTATCACCTT CTTGAAGGAC TTCACCTACA GCAAGGACGA 1260

CTTCCACCGT GCAGGCCTGC AGGTGGAGTT CATCAACCCC ATCTTCGAGT TCTCGCGGGC 1320

CATGCGGCGG CTGGGCCTGG ACGACGCTGA GTACGCCCTG CTCATCGCCA TCAACATCTT 1380

CTCGGCCGAC CGGCCCAACG TGCAGGAGCC GGGCCGCGTG GAGGCGTTGC AGCAGCCCTA 1440

CGTGGAGGCG CTGCTGTCCT ACACGCGCAT CAAGAGGCCG CAGGACCAGC TGCGCTTCCC 1500

GCGCATGCTC ATGAAGCTGG TGAGCCTGCG CACGCTGAGC TCTGTGCACT CGGAGCAGGT 1560

CTTCGCCTTG CGGCTCCAGG ACAAGAAGCT GCCGCCTCTG CTGTCGGAGA TCTGGGACGT 1620

CCACGAGTGA GGGGCTGGCC ACCCAGCCCC ACAGCCTTGC CTGACCACCC TCCAGCAGAT 1680

AGACGCCGGC ACCCCTTCCT CTTCCTAGGG TGGAAGGGGC CCTGGGCGAG CCTGTAGACC 1740

TATCGGCTCT CATCCCTTGG GATAAGCCCC AGTCCAGGTC CAGGAGGCTC CCTCCCTGCC 1800

CAGCGAGTCT TCCAGAAGGG GTGAAAGGGT TGCAGGTCCC GACCACTGAC CCTTCCCGGC 1860

TGCCCTCCCT CCCCAGCTTA CACCTCAAGC CCAGCACGCA GCGTACCTTG AACAGAGGGA 1920

GGGGAGGACC CATGGCTCTC CCCCCCTAGC CCGGGAGACC AGGGGCCTTC CTCTTCCTCT 1980

GCTTTTATTT AATAAAAATA AAAACAGAAA AAAAAAAAAA AAAAAAAAAA 2030

or a degenerate variation thereof.

8. The DNA sequence of Claim 6 comprising the sequence (SEQ ID NO:2:) which is:

Met Ser Ser Pro Thr Thr Ser Ser Leu Asp Thr Pro Leu Pro Gly Asn
1 5 10 15

Gly Pro Pro Gln Pro Gly Ala Pro Ser Ser Ser Pro Thr Val Lys Glu
20 25 30

Glu Gly Pro Glu Pro Trp Pro Gly Gly Pro Asp Pro Asp Val Pro Gly
35 40 45

Thr Asp Glu Ala Ser Ser Ala Cys Ser Thr Asp Trp Val Ile Pro Asp
50 55 60

Pro Glu Glu Glu Pro Glu Arg Lys Arg Lys Lys Gly Pro Ala Pro Lys
65 70 75 80

Met Leu Gly His Glu Leu Cys Arg Val Cys Gly Asp Lys Ala Ser Gly
85 90 95

Phe His Tyr Asn Val Leu Ser Cys Glu Gly Cys Lys Gly Phe Phe Arg
100 105 110

Arg Ser Val Val Arg Gly Gly Ala Arg Arg Tyr Ala Cys Arg Gly Gly
115 120 125

Gly Thr Cys Gln Met Asp Ala Phe Met Arg Arg Lys Cys Gln Gln Cys
130 135 140

Arg Leu Arg Lys Cys Lys Glu Ala Gly Met Arg Glu Gln Cys Val Leu
145 150 155 160

Ser Glu Glu Gln Ile Arg Lys Lys Lys Ile Arg Lys Gln Gln Gln Gln
165 170 175

Glu Ser Gln Ser Gln Ser Gln Ser Pro Val Gly Pro Gln Gly Ser Ser
180 185 190

Ser Ser Ala Ser Gly Pro Gly Ala Ser Pro Gly Gly Ser Glu Ala Gly
195 200 205

Ser Gln Gly Ser Gly Glu Gly Glu Gly Val Gln Leu Thr Ala Ala Gln
210 215 220

Glu Leu Met Ile Gln Gln Leu Val Ala Ala Gln Leu Gln Cys Asn Lys
225 230 235 240

Arg Ser Phe Ser Asp Gln Pro Lys Val Thr Pro Trp Pro Leu Gly Ala
245 250 255

Asp Pro Gln Ser Arg Asp Ala Arg Gln Gln Arg Phe Ala His Phe Thr
260 265 270

Glu Leu Ala Ile Ile Ser Val Gln Glu Ile Val Asp Phe Ala Lys Gln
275 280 285

Val Pro Gly Phe Leu Gln Leu Gly Arg Glu Asp Gln Ile Ala Leu Leu
290 295 300

Lys Ala Ser Thr Ile Glu Ile Met Leu Leu Glu Thr Ala Arg Arg Tyr
305 310 315 320

Asn His Glu Thr Glu Cys Ile Thr Phe Leu Lys Asp Phe Thr Tyr Ser
325 330 335

Lys Asp Asp Phe His Arg Ala Gly Leu Gln Val Glu Phe Ile Asn Pro
340 345 350

Ile Phe Glu Phe Ser Arg Ala Met Arg Arg Leu Gly Leu Asp Asp Ala
355 360 365

Glu Tyr Ala Leu Leu Ile Ala Ile Asn Ile Phe Ser Ala Asp Arg Pro
370 375 380

Asn Val Gln Glu Pro Gly Arg Val Glu Ala Leu Gln Gln Pro Tyr Val
385 390 395 400

Glu Ala Leu Leu Ser Tyr Thr Arg Ile Lys Arg Pro Gln Asp Gln Leu
405 410 415

Arg Phe Pro Arg Met Leu Met Lys Leu Val Ser Leu Arg Thr Leu Ser
420 425 430

Ser Val His Ser Glu Gln Val Phe Ala Leu Arg Leu Gln Asp Lys Lys
435 440 445

Leu Pro Pro Leu Leu Ser Glu Ile Trp Asp Val His Glu.
450 455 460

or a degenerate variation thereof.

9. An expression construct which comprises:

(a) a mammalian cell vector, and
(b) a base sequence encoding human steroid receptor NERI protein.

10. The expression construct of Claim 9 which comprises:

(a) vector PUC18, and
(b) a base sequence encoding human steroid receptor NERI protein.

11. The expression construct of Claim 9 wherein the base sequence comprises the sequence (SEQ ID NO:1:) which is:

CAAGAAGTGG CGAAGTTACC TTTGAGGGTA TTTGAGTAGC GGCGGTGTGT CAGGGGCTAA 60

AGAGGAGGAC GAAGAAAAGC AGAGCAAGGG AACCCAGGGC AACAGGAGTA GTTCACTCCG 120

CGAGAGGCCG TCCACGAGAC CCCCGCGCGC AGGCATGAGC CCCGCCCCCC ACGCATGAGC 180

CCCGCCCCCC GCTGTTGCTT GGAGAGgGGC GGGACCTGGA GAGAGGCTGC TCCGTGACCC 240

CACCATGTCC TCTCCTACCA CGAGTTCCCT GGATACCCCC CTGCCTGGAA ATGGCCCCCC 300

TCAGCCTGGC GCCCCTTCTT CTTCACCCAC TGTAAAGGAG GAGGGTCCGG AGCCGTGGCC 360

CGGGGGTCCG GACCCTGATG TCCCAGGCAC TGATGAGGCC AGCTCAGCCT GCAGCACAGA 420

CTGGGTCATC CCAGATCCCG AAGAGGAACC AGAGCGCAAG CGAAAGAAGG GCCCAGCCCC 480

GAAGATGCTG GGCCACGAGC TTTGCCGTGT CTGTGGGGAC AAGGCCTCCG GCTTCCACTA 540

CAACGTGCTC AGCTGCGAAG GCTGCAAGGG CTTCTTCCGG CGCAGTGTGG TCCGTGGTGG 600

GGCCAGGCGC TATGCCTGCC GGGGTGGCGG AACCTGCCAG ATGGACGCTT TCATGCGGCG 660

CAAGTGCCAG CAGTGCCGGC TGCGCAAGTG CAAGGAGGCA GGGATGAGGG AGCAGTGCGT 720

CCTTTCTGAA GAACAGATCC GGAAGAAGAA GATTCGGAAA CAGCAGCAGC AGGAGTCACA 780

GTCACAGTCG CAGTCACCTG TGGGGCCGCA GGGCAGCAGC AGCTCAGCCT CTGGGCCTGG 840

GGCTTCCCCT GGTGGATCTG AGGCAGGCAG CCAGGGCTCC GGGGAAGGCG AGGGTGTCCA 900

GCTAACAGCG GCTCAAGAAC TAATGATCCA GCAGTTGGTG GCGGCCCAAC TGCAGTGCAA 960

CAAACGCTCC TTCTCCGACC AGCCCAAAGT CACGCCCTGG CCCCTGGGCG CAGACCCCCA 1020

GTCCCGAGAT GCCCGCCAGC AACGCTTTGC CCACTTCACG GAGCTGGCCA TCATCTCAGT 1080

CCAGGAGATC GTGGACTTCG CTAAGCAAGT GCCTGGTTTC CTGCAGCTGG GCCGGGAGGA 1140

CCAGATCGCC CTCCTGAAGG CATCCACTAT CGAGATCATG CTGCTAGAGA CAGCCAGGCG 1200

CTACAACCAC GAGACAGAGT GTATCACCTT CTTGAAGGAC TTCACCTACA GCAAGGACGA 1260

CTTCCACCGT GCAGGCCTGC AGGTGGAGTT CATCAACCCC ATCTTCGAGT TCTCGCGGGC 1320

CATGCGGCGG CTGGGCCTGG ACGACGCTGA GTACGCCCTG CTCATCGCCA TCAACATCTT 1380

CTCGGCCGAC CGGCCCAACG TGCAGGAGCC GGGCCGCGTG GAGGCGTTGC AGCAGCCCTA 1440

CGTGGAGGCG CTGCTGTCCT ACACGCGCAT CAAGAGGCCG CAGGACCAGC TGCGCTTCCC 1500

GCGCATGCTC ATGAAGCTGG TGAGCCTGCG CACGCTGAGC TCTGTGCACT CGGAGCAGGT 1560

CTTCGCCTTG CGGCTCCAGG ACAAGAAGCT GCCGCCTCTG CTGTCGGAGA TCTGGGACGT 1620

CCACGAGTGA GGGGCTGGCC ACCCAGCCCC ACAGCCTTGC CTGACCACCC TCCAGCAGAT 1680

AGACGCCGGC ACCCCTTCCT CTTCCTAGGG TGGAAGGGGC CCTGGGCGAG CCTGTAGACC 1740

TATCGGCTCT CATCCCTTGG GATAAGCCCC AGTCCAGGTC CAGGAGGCTC CCTCCCTGCC 1800

CAGCGAGTCT TCCAGAAGGG GTGAAAGGGT TGCAGGTCCC GACCACTGAC CCTTCCCGGC 1860

TGCCCTCCCT CCCCAGCTTA CACCTCAAGC CCAGCACGCA GCGTACCTTG AACAGAGGGA 1920

GGGGAGGACC CATGGCTCTC CCCCCCTAGC CCGGGAGACC AGGGGCCTTC CTCTTCCTCT 1980

GCTTTTATTT AATAAAAATA AAAACAGAAA AAAAAAAAAA AAAAAAAAAA 2030

or a degenerate variation thereof.

12. COS cells transfected with the expression construct of Claim 9.

13. COS cells transfected with the expression construct of Claim 10.

14. COS cells transfected with the expression construct of Claim 11.

15. A method for determining the affinity of a test sample for a steroid hormone receptor NERI, the method comprising:

(a) constructing a chimeric gene by substituting portions of a DNA-binding domain region of a DNA sequence encoding human steroid hormone receptor NERI cDNA with operative portions of a DNA-binding domain region from a known ligand-responsive receptor protein;

(b) introducing into a suitable host cell:

(i) the chimeric gene from step (a), and

(ii) a reporter gene functionally linked to an operative hormone response element wherein the hormone response element is capable of being activated by the DNA-binding domain region of the receptor protein encoded by the chimeric gene of step (a);

(c) challenging the transfected host cell from step (b) with the test sample to be evaluated for ligand-binding-activity with the chimeric receptor protein encoded by the chimeric gene of step (a);

(d) assaying induction of the reporter gene by monitoring changes in the protein levels of the protein coded for by the reported gene.

16 The method of Claim 15 wherein the suitable host cell of step (b) is a COS cell.

17. The method of Claim 15 wherein the reporter gene of step (b)(ii) is a firefly luciferase gene.

18. The method of Claim 15 wherein the DNA sequence encoding human steroid hormone receptor NERI comprises the sequence (SEQ ID NO:1:) which is:

CAAGAAGTGG CGAAGTTACC TTTGAGGGTA TTTGAGTAGC GGCGGTGTGT CAGGGGCTAA 60

AGAGGAGGAC GAAGAAAAGC AGAGCAAGGG AACCCAGGGC AACAGGAGTA GTTCACTCCG 120

CGAGAGGCCG TCCACGAGAC CCCCGCGCGC AGGCATGAGC CCCGCCCCCC ACGCATGAGC 180

CCCGCCCCCC GCTGTTGCTT GGAGAGgGGC GGGACCTGGA GAGAGGCTGC TCCGTGACCC 240

CACCATGTCC TCTCCTACCA CGAGTTCCCT GGATACCCCC CTGCCTGGAA ATGGCCCCCC 300

TCAGCCTGGC GCCCCTTCTT CTTCACCCAC TGTAAAGGAG GAGGGTCCGG AGCCGTGGCC 360

CGGGGGTCCG GACCCTGATG TCCCAGGCAC TGATGAGGCC AGCTCAGCCT GCAGCACAGA 420

CTGGGTCATC CCAGATCCCG AAGAGGAACC AGAGCGCAAG CGAAAGAAGG GCCCAGCCCC 480

GAAGATGCTG GGCCACGAGC TTTGCCGTGT CTGTGGGGAC AAGGCCTCCG GCTTCCACTA 540

CAACGTGCTC AGCTGCGAAG GCTGCAAGGG CTTCTTCCGG CGCAGTGTGG TCCGTGGTGG 600

GGCCAGGCGC TATGCCTGCC GGGGTGGCGG AACCTGCCAG ATGGACGCTT TCATGCGGCG 660

CAAGTGCCAG CAGTGCCGGC TGCGCAAGTG CAAGGAGGCA GGGATGAGGG AGCAGTGCGT 720

CCTTTCTGAA GAACAGATCC GGAAGAAGAA GATTCGGAAA CAGCAGCAGC AGGAGTCACA 780

GTCACAGTCG CAGTCACCTG TGGGGCCGCA GGGCAGCAGC AGCTCAGCCT CTGGGCCTGG 840

GGCTTCCCCT GGTGGATCTG AGGCAGGCAG CCAGGGCTCC GGGGAAGGCG AGGGTGTCCA 900

GCTAACAGCG GCTCAAGAAC TAATGATCCA GCAGTTGGTG GCGGCCCAAC TGCAGTGCAA 960

CAAACGCTCC TTCTCCGACC AGCCCAAAGT CACGCCCTGG CCCCTGGGCG CAGACCCCCA 1020

GTCCCGAGAT GCCCGCCAGC AACGCTTTGC CCACTTCACG GAGCTGGCCA TCATCTCAGT 1080

CCAGGAGATC GTGGACTTCG CTAAGCAAGT GCCTGGTTTC CTGCAGCTGG GCCGGGAGGA 1140

CCAGATCGCC CTCCTGAAGG CATCCACTAT CGAGATCATG CTGCTAGAGA CAGCCAGGCG 1200

CTACAACCAC GAGACAGAGT GTATCACCTT CTTGAAGGAC TTCACCTACA GCAAGGACGA 1260

CTTCCACCGT GCAGGCCTGC AGGTGGAGTT CATCAACCCC ATCTTCGAGT TCTCGCGGGC 1320

CATGCGGCGG CTGGGCCTGG ACGACGCTGA GTACGCCCTG CTCATCGCCA TCAACATCTT 1380

CTCGGCCGAC CGGCCCAACG TGCAGGAGCC GGGCCGCGTG GAGGCGTTGC AGCAGCCCTA 1440

CGTGGAGGCG CTGCTGTCCT ACACGCGCAT CAAGAGGCCG CAGGACCAGC TGCGCTTCCC 1500

GCGCATGCTC ATGAAGCTGG TGAGCCTGCG CACGCTGAGC TCTGTGCACT CGGAGCAGGT 1560

CTTCGCCTTG CGGCTCCAGG ACAAGAAGCT GCCGCCTCTG CTGTCGGAGA TCTGGGACGT 1620

CCACGAGTGA GGGGCTGGCC ACCCAGCCCC ACAGCCTTGC CTGACCACCC TCCAGCAGAT 1680

AGACGCCGGC ACCCCTTCCT CTTCCTAGGG TGGAAGGGGC CCTGGGCGAG CCTGTAGACC 1740

TATCGGCTCT CATCCCTTGG GATAAGCCCC AGTCCAGGTC CAGGAGGCTC CCTCCCTGCC 1800

CAGCGAGTCT TCCAGAAGGG GTGAAAGGGT TGCAGGTCCC GACCACTGAC CCTTCCCGGC 1860

TGCCCTCCCT CCCCAGCTTA CACCTCAAGC CCAGCACGCA GCGTACCTTG AACAGAGGGA 1920

GGGGAGGACC CATGGCTCTC CCCCCCTAGC CCGGGAGACC AGGGGCCTTC CTCTTCCTCT 1980

GCTTTTATTT AATAAAAATA AAAACAGAAA AAAAAAAAAA AAAAAAAAAA 2030

or a degenerate variation thereof.

```
1     CAAGAAGTGGCGAAGTTACCTTTGAGGGTATTTGAGTAGCGGCGGTGTGTCAGGGGCTAA
61    AGAGGAGGACGAAGAAAAGCAGAGCAAGGGAACCCAGGGCAACAGGAGTAGTTCACTCCG
121   CGAGAGGCCGTCCACGAGACCCCCGCGCGCAGGCATGAGCCCCGCCCCCCACGCATGAGC
181   CCCGCCCCCCCGCTGTTGCTTGGAGAGGGGCGGGACCTGGAGAGAGGCTGCTCCGTGACCC
241   CACCATGTCCTCTCCTACCACGAGTTCCCTGGATACCCCCCTGCCTGGAAATGGCCCCCC
          M  S  S  P  T  T  S  S  L  D  T  P  L  P  G  N  G  P  P   19
301   TCAGCCTGGCGCCCCTTCTTCTTCACCCACTGTAAAGGAGGAGGGTCCGGAGCCGTGGCC
      Q  P  G  A  P  S  S  S  P  T  V  K  E  E  G  P  E  P  W  P   39
361   CGGGGGTCCGGACCCTGATGTCCCAGGCACTGATGAGGCCAGCTCAGCCTGCAGCACAGA
      G  G  P  D  P  D  V  P  G  T  D  E  A  S  S  A  C  S  T  D   59
421   CTGGGTCATCCCAGATCCCGAAGAGGAACCAGAGCGCAAGCGAAAGAAGGGCCCAGCCCC
      W  V  I  P  D  P  E  E  E  P  E  R  K  R̃  K  K  G  P  A  P   79
481   GAAGATGCTGGGCCACGAGCTTTGCCGTGTCTGTGGGGACAAGGCCTCCGGCTTCCACTA
      K  M  L  G  H  E  L  C  R  V  C  G  D  K  A  S  G  F  H  Y   99
541   CAACGTGCTCAGCTGCGAAGGCTGCAAGGGCTTCTTCCGGCGCAGTGTGGTCCGTGGTGG
      N  V  L  S  C  E  G  C  K  G  F  F  R  R  S  V  V  R  G  G  119
601   GGCCAGGCGCTATGCCTGCCGGGGTGGCGGAACCTGCCAGATGGACGCTTTCATGCGGCG
      A  R  R  Y  A  C  R  G  G  G  T  C  Q  M  D  A  F  M  R  R  139
661   CAAGTGCCAGCAGTGCCGGCTGCGCAAGTGCAAGGAGGCAGGGATGAGGGAGCAGTGCGT
      K  C  Q  Q  C  R  L  R  K  C  K  E  A  G  M  R  E  Q  C  V  159
721   CCTTTCTGAAGAACAGATCCGGAAGAAGAAGATTCGGAAACAGCAGCAGCAGGAGTCACA
      L  S  E  E  Q  I  R  K  K  K  I  R  K  Q  Q  Q  Q  E  S  Q  179
781   GTCACAGTCGCAGTCACCTGTGGGGCCGCAGGGCAGCAGCAGCTCAGCCTCTGGGCCTGG
      S  Q  S  Q  S  P  V  G  P  Q  G  S  S  S  S  A  S  G  P  G  199
841   GGCTTCCCCTGGTGGATCTGAGGCAGGCAGCCAGGGCTCCGGGGAAGGCGAGGGTGTCCA
      A  S  P  G  G  S  E  A  G  S  Q  G  S  G  E  G  E  G  V  Q  219
901   GCTAACAGCGGCTCAAGAACTAATGATCCAGCAGTTGGTGGCGGCCCAACTGCAGTGCAA
      L  T  A  A  Q  E  L  M  I  Q  Q  L  V  A  A  Q  L  Q  C  N  239
961   CAAACGCTCCTTCTCCGACCAGCCCAAAGTCACGCCCTGGCCCCTGGGCGCAGACCCCCA
      K  R  S  F  S  D  Q  P  K  V  T  P  W  P  L  G  A  D  P  Q  259
1021  GTCCCGAGATGCCCGCCAGCAACGCTTTGCCCACTTCACGGAGCTGGCCATCATCTCAGT
      S  R  D  A  R  Q  Q  R  F  A  H  F  T  E  L  A  I  I  S  V  279
```

FIG.1

```
1081  CCAGGAGATCGTGGACTTCGCTAAGCAAGTGCCTGGTTTCCTGCAGCTGGGCCGGGAGGA
        Q  E  I  V  D  F  A  K  Q  V  P  G  F  L  Q  L  G  R  E  D   299
1141  CCAGATCGCCCTCCTGAAGGCATCCACTATCGAGATCATGCTGCTAGAGACAGCCAGGCG
        Q  I  A  L  L  K  A  S  T  I  E  I  M  L  L  E  T  A  R  R   319
1201  CTACAACCACGAGACAGAGTGTATCACCTTCTTGAAGGACTTCACCTACAGCAAGGACGA
        Y  N  H  E  T  E  C  I  T  F  L  K  D  F  T  Y  S  K  D  D   339
1261  CTTCCACCGTGCAGGCCTGCAGGTGGAGTTCATCAACCCCATCTTCGAGTTCTCGCGGGC
        F  H  R  A  G  L  Q  V  E  F  I  N  P  I  F  E  F  S  R  A   359
1321  CATGCCGGCGCTGGGCCTGGACGACGCTGAGTACGCCCTGCTCATCGCCATCAACATCTT
        M  R  R  L  G  L  D  D  A  E  Y  A  L  L  I  A  I  N  I  F   379
1381  CTCGGCCGACCGCCCAACGTGCAGGAGCCGGGCCGCGTGGAGGCGTTGCAGCAGCCCTA
        S  A  D  R  P  N  V  Q  E  P  G  R  V  E  A  L  Q  Q  P  Y   399
1441  CGTGGAGGCGCTGCTGTCCTACACGCGCATCAAGAGGCCGCAGGACCAGCTGCGCTTCCC
        V  E  A  L  L  S  Y  T  R  I  K  R  P  Q  D  Q  L  R  F  P   419
1501  GCGCATGCTCATGAAGCTGGTGAGCCTGCGCACGCTGAGCTCTGTGCACTCGGAGCAGGT
        R  M  L  M  K  L  V  S  L  R  T  L  S  S  V  H  S  E  Q  V   439
1561  CTTCGCCTTGCGGCTCCAGGACAAGAAGCTGCCGCCTCTGCTGTCGGAGATCTGGGACGT
        F  A  L  R  L  Q  D  K  K  L  P  P  L  L  S  E  I  W  D  V   459
1621  CCACGAGTGAGGGGCTGGCCACCCAGCCCCACAGCCTTGCCTGACCACCCTCCAGCAGAT
        H  E  *  461
1681  AGACGCCGGCACCCCTTCCTCTTCCTAGGGTGGAAGGGGCCCTGGGCGAGCCTGTAGACC
1741  TATCGGCTCTCATCCCTTGGGATAAGCCCCAGTCCAGGTCCAGGAGGCTCCCTCCCTGCC
1801  CAGCGAGTCTTCCAGAAGGGGTGAAAGGGTTGCAGGTCCCGACCACTGACCCTTCCCCGGC
1861  TGCCCTCCCTCCCCAGCTTACACCTCAAGCCCAGCACGCAGCGTACCTTGAACAGAGGGA
1921  GGGGAGGACCCATGGCTCTCCCCCCCTAGCCCGGGAGACCAGGGGCCTTCCTCTTCCTCT
1981  GCTTTTATTTAATAAAAATAAAAACAGAAAAAAAAAAAAAAAAAAAAAA  2030
```

FIG.1(Cont.)

```
SENSE                  C   E   G   C   K   A/G
ES11       5' CGAATTC TGT GAG GGC TGC AAG GCC
                                       A   G

SENSE                  C   E   G   C   K   A/G F   F
ES12       5' CGAATTC TGT GAG GGC TGC AAG GCC TTC TTC
                                       A   G

ANTISENSE              F   G   P   X   X   K   A   F
ES15       5' GGAATTC AAA ACC AGG IAI III CTT TGC AAA
                       G   C   C           T   A   G
                           T   T               G
                           G   G               C
```

FIG.2A

```
5' GAGGGAGCAGT·GCGTCCTTT·CTGAAGAA Ner
   || | |||  | ||  || |  | |||
3' CTTCTTCGGGAACGTCGGGAGTGTCTTAGC E12
```

FIG.2B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/09165

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C07K 13/00; A61K 37/24, 37/26; C12P 21/06
US CL :530/399; 536/23.1; 435/69.4, 69.7
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/399; 536/23.1; 435/69.4, 69.7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Dialog, Intelligenetics

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Science, Vol. 240, issued 13 May 1988, R.M. Evans, "The steroid and thyroid hormone receptor family", pages 889-895.	1-18
A	US, A, 4,981,784 (Evans et al.) 01 January 1991.	1-18
A	Nature, Vol. 332, issued 28 April 1988, N. Brand et al., "Identification of a second human retinoic acid receptor", pages 850-853.	1-18

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
17 November 1993

Date of mailing of the international search report
08 DEC 1993

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. NOT APPLICABLE

Authorized officer
KAREN COCHRANE CARLSON, PH.D.

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*